PCT/AU03/01142

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Patent Office
Canberra

I, JONNE YABSLEY, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. 2002951224 for a patent by JOHNSON & JOHNSON RESEARCH PTY LTD as filed on 04 September 2002.

CERTIFIED COPY OF PRIORITY DOCUMENT

WITNESS my hand this
Twelfth day of September 2003

JONNE YABSLEY
TEAM LEADER EXAMINATION
SUPPORT AND SALES

AUSTRALIA

Patents Act 1990

Johnson & Johnson Research Pty Ltd

PROVISIONAL SPECIFICATION

Invention Title:

RNAi Library

The invention is described in the following statement:

141360837

RNAi LIBRARY

FIELD OF THE INVENTION

The present invention relates to an RNAi library. More particularly the invention relates to the construction of a gene expression library encoding RNA capable of suppressing an expressed gene.

BACKGROUND OF THE INVENTION

The introduction of double stranded RNA (dsRNA) into a range of organisms induces both a potent and specific gene silencing effect. This form of gene suppression by a dsRNA molecule was first observed in *Caenorhabditis elegans* and given the term RNA interference or RNAi (Fire et al 1998). In an attempt to optimise the use of antisense RNA as a tool for controlling specific gene expression in worms, Fire et al (1998) found that dsRNA was more effective than antisense RNA alone. The dsRNA could be generated *in vitro* (Fire et al 1998) or *in vivo* (Tavernarakis et al 2000) and still mediate gene suppression with high specificity. Subsequent studies have shown that dsRNA is an effective inducer of gene silencing in a wide range of eukaryotic organisms and that the mechanism behind this form of gene regulation is most likely conserved throughout evolution (Baulcombe, D. C. (1996) *Plant Mol Biol* 32(1-2), 79-88; Lohmann, J. U., Endl, I., and Bosch, T. C. (1999) *Dev Biol* 214(1), 211-4; Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998) *Proc Natl Acad Sci U S A* 95(25), 14687-92; Cogoni, C., and Macino, G. (1999) *Nature* 399(6732), 166-9; Kennerdell, J. R., and Carthew, R. W. (1998) *Cell* 95(7), 1017-26; Schoppmeier, M., and Damen, W. G. (2001) *Dev Genes Evol* 211(2), 76-82; Baker, M. W., and Macagno, E. R. (2000) *Curr Biol* 10(17), 1071-4; Wargelius, A., Ellingsen, S., and Fjose, A. (1999) *Biochem Biophys Res Commun* 263(1), 156-61).

The molecular mechanism of RNAi has begun to be deciphered using biochemical and genetic approaches in different experimental systems (Hammond, S.M., Caudy, A.A., and Hannon, G.J. (2001) *Nat. Rev. Genet.* 2, 110-19). Presently, RNAi is postulated to involve both an initiation step and an effector step. During the initiation phase, dsRNA is processed by the RNaseIII family nuclease Dicer to produce 21-23 nucleotide duplex siRNAs (small interfering RNAs). These short stretches of dsRNA carry 2 nucleotide 3'-OH overhangs that contribute to the efficacy of gene

silencing (Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001) *Genes & Dev* 15:188-200). In the effector phase, these siRNAs are incorporated into a multiprotein complex called RISC (RNA-induced silencing complex) that targets transcripts by base pairing between one of the siRNA strands and the endogenous mRNA (Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000) *Nature* 404: 293-96). A nuclease activity associated with the RISC complex then cleaves the mRNA-siRNA duplex thus targeting the cognate mRNA for destruction.

In mammalian cells the use of dsRNA to control gene expression has been hampered by the presence of a unique global response mechanism. Mammalian cells exposed to dsRNA longer than 30 base pairs in length trigger a response mechanism involving activation of two key enzymes, dsRNA-activated protein kinase (PKR) and 2'5' oligoadenylate polymerase/RnaseL (Kumar, M., and Carmichael, G. G. (1998) *Microbiol Mol Biol Rev* 62(4), 1415-34). The activation of these enzymes leads to a cessation of protein synthesis and eventually cell death via apoptosis. It was thus anticipated that the introduction of long dsRNA would activate this global response system. However, studies have shown that in both mouse pre-implantation embryos (Svoboda, P., Stein, P., Hayashi, H., and Schultz, R. M. (2000) *Development* 127(19), 4147-4156; Wianny, F., and Zernicka-Goetz, M. (2000) *Nat Cell Biol* 2(2), 70-5) and undifferentiated embryonic stem cells and embryonic carcinoma cells (Yang, S., Tutton, S., Pierce, E., and Yoon, K. (2001) *Mol Cell Biol* 21(22), 7807-16; Billy, E., Brondani, V., Zhang, H., Muller, U. and Filipowicz, W. (2001) *Proc. Natl Acad Sci* 98, 14428-14483; Paddison, P., Caudy, A. A., and Hannon, G.J. (2002) *Proc. Natl Acad. Sci.* 99, 1443-1448), the use of *in vitro* generated long dsRNA was able to mediate specific gene silencing. The primary reason for these observations was that these cell systems lack the generalised responses to dsRNA. These results were encouraging but placed particular limitations on the utility of this approach in differentiated mammalian cells.

Following on from observations that the products of the Dicer enzyme could mediate RNAi in Drosophila embryo extracts, it was then demonstrated that chemically synthesised 21 bp siRNAs could be used in a wide range of human and mouse cell lines to induce gene silencing (Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001) *Nature* 411(6836), 494-8; Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001) *Proc. Natl. Acad. Sci.* 98, 9742-9747). This approach for transiently controlling the expression of a wide range of different target genes has been demonstrated and is becoming the method of choice for determining gene function in mammalian cells (Hsu, J.Y., Reimann, J. D. R.,

Sorensen, C.S., Lucas, J., and Jackson, P. K. (2002) *Nature Cell Biol.* 4, 358-366; Thompson, B., Tonwsley, F., Rosin-Arbesfeld, R., Muisi, H., and Bienz, M. (2002) *Nature Cell Biol.* 4, 367-373). One of the limitations associated with these synthetic dsRNA strategies is the transient nature of the suppressive effect induced by the dsRNA.

More recently, it has been shown that mammalian cells contain a very large group of small RNAs called microRNAs which are postulated to be transcribed as hairpin RNA precursors that are processed by Dicer to produce the mature 21 base forms (Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001) *Science* 294, 853-858; Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001) *Science* 294, 858-862; Lee, R.C. and Ambros, V. (2001) *Science* 294, 862-864). Several groups have exploited this naturally occurring biological mechanism to show that short hairpin RNAs (shRNAs) can induce specific gene silencing in mammalian cells (Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002) *Genes & Dev* 16, 948-958; Brummelkamp, T.R., Bernards, R., and Agami, R. (2002) *Science* 296, 550-553; Sui, G., Soohoo, C., Affar, E., Gay, F., Shi, Y., Forrester, W. C., and Shi, Y. (2002) *Proc Natl Acad Sci* 99, 5515-20; Yu, J., DeRuiter, S.L., and Turner, D.L. (2002) *Proc Natl Acad Sci* 99, 6047-52). Furthermore, expression cassettes have been developed using the endogenous U6 snRNA or H1 promoters to drive expression of sequence-specific shRNAs that can regulate gene expression both transiently and stably in mammalian cells via RNAi (Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002) *Genes & Dev* 16, 948-958; Brummelkamp, T.R., Bernards, R., and Agami, R. (2002) *Science* 296:550-553). ShRNAs produced from these expression cassettes were processed by Dicer to 21 bp siRNAs which are believed to be the effectors of gene silencing. It is anticipated that these cassettes will be useful for reverse genetic approaches in mammalian cells and transgenic mice to better understand gene function, and also as therapeutics.

A major limitation with the state of the art for RNAi in mammalian cells is the lack of any strategy for using RNAi knockdowns in a forward genetic approach to identify new genes involved in cellular processes or different human diseases. At present, synthetic siRNAs or RNAi expression constructs are designed on a gene-by-gene basis limiting the utility of these strategies for both generating and screening genome-wide RNAi expression libraries. The present invention provides methods which enable the production of RNAi libraries.

SUMMARY OF THE INVENTION

In a first aspect the present invention provides a method of producing a DNA molecule wherein mRNA transcribed from the DNA molecule forms hairpin RNA (hRNA), the method comprising:

synthesizing a first DNA strand comprising in order a first sequence, a random sequence and a second sequence, wherein the second sequence has a 3' end and a 5' end such that nucleotides at the 3' end of the second sequence are complementary to nucleotides at the 5' end of the second sequence such that the second sequence forms a stem loop;

synthesizing a complementary DNA strand extending from the stem loop using a DNA polymerase, the complementary DNA strand being complementary to the first region and the random sequence so as to form hairpin DNA;

denaturing the hairpin DNA to form a single DNA strand;

adding a primer having the same sequence as the first region and DNA polymerase to synthesize double stranded DNA.

In a second aspect the present invention provides a method of preparing an expression vector, expression of which produces dsRNA, the method comprising:

synthesizing a first DNA strand comprising in order at least four consecutive adenosine nucleotides, a random sequence, at least four consecutive thymidine nucleotides and a primer binding site;

annealing a primer to the primer binding site and synthesizing a second DNA strand which is substantially complementary to the first DNA strand and forms double stranded DNA; and

cloning the double stranded DNA into an expression vector between two convergent promoters.

In a third aspect the present invention provides a method for determining a function of a gene, the method comprising:

synthesizing a first DNA strand comprising in order a first sequence, a random sequence and a second sequence, wherein the second sequence has a 3' end and a 5' end such that nucleotides at the 3' end of the second sequence are complementary to nucleotides at the 5' end of the second sequence such that the second sequence forms a stem loop;

synthesizing a complementary DNA strand extending from the stem loop using a DNA polymerase, the complementary DNA strand being complementary to the first region and the random sequence so as to form hairpin DNA;

denaturing the hairpin DNA to form a single DNA strand;

adding a primer having the same sequence as the first region and DNA polymerase to synthesize double stranded DNA;

cloning the double stranded DNA into an expression vector wherein the double stranded DNA is under the control of a promoter;

transfecting an effective amount of the expression vector into a cell under conditions permitting transcription of the double stranded DNA to produce a transfected cell;

detecting one or more changes selected from morphological changes and behavioural changes in the transfected cell relative to a control cell.

In a fourth aspect the present invention provides a method for determining a function of a gene, the method comprising:

synthesizing a first DNA strand comprising in order at least four consecutive adenosine nucleotides, a random sequence, at least four consecutive thymidine nucleotides and a primer binding site;

annealing a primer to the primer binding site and synthesizing a second DNA strand which is substantially complementary to the first DNA strand and forms double stranded DNA; and

cloning the double stranded DNA into an expression vector between two convergent promoters;

transfecting an effective amount of the expression vector into a cell under conditions favouring transcription of the double stranded DNA to produce a transfected cell;

detecting one or more changes selected from morphological changes and behavioural changes in the transfected cell relative to a control cell.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 Generation of a random shRNA library.

Fig. 2 Generation of a random siRNA library.

Fig. 3 Method for constructing genome-specific shRNA and siRNA libraries.

Fig. 4 Identification of HIV-specific shRNA or siRNA using genetic selections.

Fig.5 p53 shRNA construction scheme

DETAILED DESCRIPTION

The present invention relates to methods which enable production of a library of shRNA or siRNA genes encoding shRNA or siRNAs that are capable of recognising all target mRNA sites to identify, isolate and characterise unknown and known genes that contribute to a specific cellular phenotype or are modified by specific stimuli. These expression libraries are designed to suppress the expression of a target gene and based on the sequence of the encoded shRNA or siRNA identify the target gene responsible for the change in cellular phenotype. This method requires the construction of random shRNA and siRNA expression libraries that contain inserts encoding RNA sequences that form double-stranded RNA via intramolecular or intermolecular hybridisation in vivo, respectively.

The use of dsRNA as a mediator has distinct advantages over hammerhead and hairpin ribozymes including the presence of a natural cellular protein complex (termed RISC) that binds the expressed dsRNA and mediates interaction with the target mRNA and cleavage of the target mRNA.

In a first aspect the present invention provides a method of producing a DNA molecule wherein mRNA transcribed from the DNA molecule forms hairpin RNA (hRNA), the method comprising:

synthesizing a first DNA strand comprising in order a first sequence, a random sequence and a second sequence, wherein the second sequence has a 3' end and a 5' end such that nucleotides at the 3' end of the second sequence are complementary to nucleotides at the 5' end of the second sequence such that the second sequence forms a stem loop;

synthesizing a complementary DNA strand extending from the stem loop using a DNA polymerase, the complementary DNA strand being complementary to the first region and the random sequence so as to form hairpin DNA;

denaturing the hairpin DNA to form a single DNA strand;

adding a primer having the same sequence as the first region and DNA polymerase to synthesize double stranded DNA.

In a preferred embodiment a deoxyuracil nucleotide is included in the first sequence and prior to addition of the primer the single DNA strand is depurinated with uracil nucleotide glycosylase and β-elimated.

In a preferred embodiment the double stranded DNA is cloned into an expression vector. More preferably the double stranded DNA is cloned into an expression vector wherein the double stranded DNA is under the control of a promoter.

In a preferred embodiment the first DNA strand comprises a restriction enzyme site.

Delivery and transcription of the expression vectors of the present invention in a host cell provides a shRNA specific for a target mRNA having complementarity with the double-stranded RNA region. The shRNAs of the invention have been shown to be effective modifiers of gene expression.

Preferably the random sequence is about 19 to about 30 base pairs in length. More preferably the random sequence is from 19 to 25 base pairs in length. Most preferably the random sequence is 19 base pairs in length.

The random sequence may be synthetically generated by random insertion of nucleotides during synthesis or by randomly digesting a viral or pathogenic genome to fragment the starting DNA. Random digestion of a viral or pathogenic genome may be achieved by techniques known to those of skill in the art, such as DNAse I digestion. Synthetic sequences may be generated chemically according to known methods such as the solid phase phosphoramidite triester method described by Beaucage and Caruthers (1981) Tetrahedron Letts. 22(20):1859-1862, e.g. using an automated synthesiser as described in Needham-VanDevanter et al (1984) Nucleic Acids Res., 12:6159-6168. Purification of the molecule, where necessary, is typically performed by either gel electrophoresis or by anion-exchange HPLC as described in Pearson and Regnier (1983) J. Chrom. 255:137-149. The sequence can be verified using the chemical degradation method of Maxam and Gilbert (1980) in Grossman and Moldave (eds.) Academic Press, New York, Methods in Enzymology 65:499-560.

As used herein, the term "complementary" is used in reference to "polynucleotides" and oligonucleotides" (which are interchangeable terms that refer to a sequence of nucleotides) related by the base pairing rules. For example, the

sequence 5'-CTGAG-3' is complementary to the sequence 5'- CTCAG-3'. Complementarity can be partial or total. Partial complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. Total or complete complementarity is where each and every nucleic acid base is matched with another base according to base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridisation between nucleic acid strands.

The term "loop" refers to an unpaired secondary structure in an RNA sequence in which a single-stranded RNA sequence is flanked by RNA sequences which are capable of pairing with each other to form a "stem" structure. The term "unpaired" when made in reference to RNA refers to a secondary structure in an RNA sequence in which RNA is single-stranded and is flanked by RNA sequences which are incapable of pairing with each other, but which are capable of pairing with other sequences. Loop structures of any length and any sequence are contemplated to be within the scope of this invention. Computer programs for the prediction of RNA secondary structure formation are known in the art and include, for example, "RNAFOLD" described in Hofacker et al. (1994) Monatshefte F. Chemie 125:167-188; McCaskill (1990) Biopolymers 29:1105-1119 and "DNASIS" (Hitachi).

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a host organism. Nucleic acid sequences necessary for expression in eukaryotic cells usually include a promoter and termination and polyadenylation signals. In a preferred embodiment the expression vector also incorporates stabilisation elements into the expressed RNA to increase the stability of the RNA. As used herein, the term "vector" is used in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another. Vector includes plasmids, viruses, retrotransposons and cosmids.

Preferably the double stranded DNA is cloned into an expression vector suitable for expression in a mammalian cell. Methods which are well known to those skilled in the art can be used to construct expression vectors containing a sequence which encodes the RNA expression library. These methods include in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination or genetic recombination. Such techniques are described in Sambrook et al (1989) Molecular Cloning, A laboratory Manual, Cold Spring Harbor Press, Plainview N.Y.

and Asubel F M et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York N.Y.

As used herein, the term "promoter" refers to a single promoter sequence as well as to a plurality (i.e., one or more) of promoter sequences which are operably linked to each other and to at least one DNA sequence of interest. Promoters consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis T. et al., Science 236:1237 (1987). Promoter elements have been isolated from a variety of eukaryotic sources including genes in plant, yeast, insect and mammalian cells and viruses. The selection of a particular promoter depends on what cell type is to be used to express the DNA sequence of interest. The promoter also optionally includes distal enhancer or repressor elements which can be located as much as several thousand base pairs from the site of transcription. The promoter may be constitutive, such as a promoter active under most environmental conditions or stages of development or the promoter may be inducible, and respond to, for example, an extracellular stimulus.

Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are generally a few hundred nucleotides in length.

In a preferred embodiment the double stranded DNA is cloned into an expression vector under the control of a U6 snRNA, H1 or T7 promoter. More preferably the double stranded DNA is cloned in to an expression vector under the control of a U6 snRNA promoter.

Synthesis of the second DNA strand may be achieved using second strand synthesis techniques well known to those of skill in the art for synthesizing a second strand of DNA from a first strand of DNA, for example utilizing a DNA polymerase such as AmpliTaq DNA polymerase (Perkin Elmer). Suitable techniques for second strand synthesis may be as set out in Sambrook et al (1989) Molecular Cloning, A laboratory Manual, Cold Spring Harbor Press, Plainview N.Y. and Asubel F M et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York N.Y.

In a second aspect the present invention provides a method of preparing an expression vector, expression of which produces dsRNA, the method comprising:

synthesizing a first DNA strand comprising in order at least four consecutive adenosine nucleotides, a random sequence, at least four consecutive thymidine nucleotides and a primer binding site;

annealing a primer to the primer binding site and synthesizing a second DNA strand which is substantially complementary to the first DNA strand and forms double stranded DNA; and

cloning the double stranded DNA into an expression vector between two convergent promoters.

Transcription from the convergent promoters of two strands of the resident inserts results in the production of two small complementary RNAs that are capable of hybridising to form an siRNA with two to four base overhangs at their 3' ends.

The expression vector produced according to the methods of the invention are useful in identifying the function of a gene or sequence of interest in an organism.

Preferably the random sequence is about 19 to about 30 base pairs in length. More preferably the random sequence is from 19 to 25 base pairs in length. Most preferably the random sequence is 19 base pairs in length.

In a preferred embodiment the double stranded DNA is cloned in to an expression vector between two convergent U6 snRNA, H1 or T7 promoters. More preferably the double stranded DNA is cloned in to an expression vector between two convergent U6 snRNA promoters.

In a preferred embodiment, the expression vectors prepared according to the methods of the first or second aspect are used to transfect a host cell.

In a third aspect the present invention provides a method for determining a function of a gene, the method comprising:

synthesizing a first DNA strand comprising in order a first sequence, a random sequence and a second sequence, wherein the second sequence has a 3' end and a 5' end such that nucleotides at the 3' end of the second sequence are complementary to nucleotides at the 5' end of the second sequence such that the second sequence forms a stem loop;

synthesizing a complementary DNA strand extending from the stem loop using a DNA polymerase, the complementary DNA strand being complementary to the first region and the random sequence so as to form hairpin DNA;

denaturing the hairpin DNA to form a single DNA strand;

adding a primer having the same sequence as the first region and DNA polymerase to synthesize double stranded DNA;

cloning the double stranded DNA into an expression vector wherein the double stranded DNA is under the control of a promoter;

transfecting an effective amount of the expression vector into a cell under conditions permitting transcription of the double stranded DNA to produce a transfected cell;

detecting one or more changes selected from morphological changes and behavioural changes in the transfected cell relative to a control cell.

In a fourth aspect the present invention provides a method for determining a function of a gene, the method comprising:

synthesizing a first DNA strand comprising in order at least four consecutive adenosine nucleotides, a random sequence, at least four consecutive thymidine nucleotides and a primer binding site;

annealing a primer to the primer binding site and synthesizing a second DNA strand which is substantially complementary to the first DNA strand and forms double stranded DNA; and

cloning the double stranded DNA into an expression vector between two convergent promoters;

transfecting an effective amount of the expression vector into a cell under conditions favouring transcription of the double stranded DNA to produce a transfected cell;

detecting one or more changes selected from morphological changes and behavioural changes in the transfected cell relative to a control cell.

The present invention provides methods for the identification of one or more functions of a nucleotide sequence in an organism. The methods of the invention selectively reduce, diminish or destroy the RNA encoded by the targeted coding sequence in order to render the RNA non-functional while the targeted gene in the host remains intact. These methods therefore employ a "knockdown" strategy to determine gene function instead of the traditional "knockout" methods. The invention is useful for the rapid identification of, for example, disease related genes which may be targeted for the treatment or prevention of disease. The methods of the present invention also have utility in identifying viral or pathogen-derived genes that play a major role in the susceptibility of cells to infection by viruses or pathogens.

In a preferred embodiment the expression vector is a retroviral expression vector.

In a preferred embodiment the transfected cell is recovered and the double stranded DNA insert recovered or amplified by, for example, using the polymerase chain reaction, re-cloned and subjected to additional enrichment steps.

In a further preferred embodiment the enriched insert is sequenced and used to identify potential target genes by, for example, homology searching, or utilised to capture the target mRNA.

In a preferred embodiment the expression vector encodes a selectable marker, for example an antibiotic resistance gene, for selection of cells transfected with the expression vector. More preferably the expression vector encodes the G418 selection marker.

The term "transfection" as used herein refers to the introduction of a transgene, for example a vector, into a cell. Transfection may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, retroviral infection or biolistics (i.e., particle bombardment). Transfection may be transient or stable transfection. The term "stable transfection" or "stably transfected" refers to the introduction and integration of a transgene into the genome of a transfected cell. The term "transient transfection" or "transiently transfected" refers to the introduction of one or more transgenes into a transfected cell in the absence of integration of the transgene into the genome of the host cell.

The term "gene of interest" refers to any gene , the manipulation of which may be deemed desirable for any reason, by one of ordinary skill in the art.

In a preferred embodiment the methods of the present invention for determining the function of a genomic DNA sequence, a shRNA or siRNA sequence is introduced into a cell in order to reduce the amount of RNA expressed by that genomic sequence.

It is desirable to express a sufficient amount of shRNA or siRNA such that substantially all the substrate RNA is cleaved. Such substantial abrogation of substrate RNA expression would facilitate the observation of the effect of depletion of gene function in the organism wherein the shRNA or siRNA is expressed. While desirable, complete elimination of the substrate RNA is not required by the methods of the invention.

A "control" cell as used herein includes a cell that is untransfected, has been mock transfected, or has been transfected with an "empty vector" such as an expression vector without the double stranded DNA insert.

Host cells, such as eukaryotic cells, harbouring the expression vectors described above also are provided by this invention. Suitable host cells include, but are not limited to, bacterial cells, rat cells, mouse cells and human cells.

The methods of the invention are useful for determining the function of a gene or DNA sequence of interest in an organism by forward genetic approaches including observing the effects of reducing expression of the gene or DNA sequence in the organism or of a homologous gene or DNA sequence in another organism. For example, data presented herein demonstrates that the function of the p53 or EGFP gene in HCT116 colon cancer cells or HEK 293 embryonic kidney cells respectively may be determined by siRNA or shRNA mediated cleavage of transcripts.

The types of genetic selections that can be used in a forward genetic approach with a genome-wide RNAi library includes overcoming cell growth arrest by, for example, bypassing p53-mediated growth arrest and apoptosis; identifying new targets involved in chemotherapeutic drug resistance such as overcoming 5-FU-induced growth arrest, apoptosis and senescence; blocking activated signaling pathways, for example, identifying novel positive and negative regulators of signaling pathways implicated in cancer, such as the TGFβ and Wnt pathways; elucidating resistance to viral and pathogen infection including genetic screens for genes that confer resistance to HIV infection or that interfere with the productive or latent phases of the viral life cycle or genetic screens for genes that interfere with the lifecycle of an intracellular parasite such as plasmodium; synthetic lethality screens to identify gene products whose inactivation leads to cell death, particularly in tumor cells deficient for either the p53 or p16/Rb tumor suppression pathways; identifying genes involved in metastasis, for example using in vivo assays; identifying optimal siRNAs against specific target(s); detecting genes regulating specific promoters; detecting cell cycle regulatory genes, for example using soft agar assays (for anchorage dependent growth) and minimal medium (for growth factor-independent growth), both of which are widely used indicators of cellular transformation in cell culture; identifying unknown genes responsible for tumorigenesis such as using bromo-deoxyuridine, a nucleoside analog that is toxic to cells undergoing active division.

The function of a gene of interest in a first organism may be determined by reducing expression of a gene in a second organism using the methods of the

invention where ethical considerations preclude experimentation on the first organism of interest. Thus, while ethical considerations preclude depleting gene expression in humans in order to determine the function of that gene in a human individual, the ability to reduce expression of a gene sequence that is homologous to the human gene sequence in a model organism (e.g. mouse) permits an initial determination of the function of the gene. The function of the gene may then be further investigated in other model organisms and/or in clinical trials with human subjects.

Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

All publications mentioned in this specification are herein incorporated by reference. Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed in Australia before the priority date of each claim of this application.

In order that the nature of the present invention may be more clearly understood preferred forms thereof will now be described with reference to the following non-limiting Examples.

EXAMPLES

Example 1 Generation of random shRNA expression libraries.

With the current state of the art, gene-expressed shRNAs are designed and constructed on a gene-by-gene basis (Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002) *Genes & Dev* 16, 948-958; Brummelkamp, T.R., Bernards, R., and Agami, R. (2002) *Science* 296, 550-553; Sui, G., Soohoo, C., Affar, E., Gay, F., Shi, Y., Forrester, W. C., and Shi, Y. (2002) *Proc Natl Acad Sci* 99, 5515-20; Yu, J., DeRuiter, S.L., and Turner, D.L. (2002) *Proc Natl Acad Sci* 99, 6047-52). To adapt the shRNA strategy for utility against all expressed sequences in a mammalian cell, the applicants have designed a method for constructing a universal random shRNA expression library. The steps involved are outlined in Figure 1. Initially, a pool of single-stranded DNA is synthesised containing a primer-binding site, a restriction

enzyme site (RE1), a 19 base random sequence (N19) and a stem loop (Figure 1A). The design of this stem loop region is such that the 3' end of the oligonucleotide bases pairs with complementary bases 5' to the loop sequence. This results in a substrate that can undergo self-priming. Using this feature of the oligonucleotide pool, DNA polymerase is added to synthesise the complementary strands for both the random sequence (C-19) and the primer-binding site (Figure 1B). Following denaturation of these structures, a primer is added that is the same sequence as the primer-binding site and therefore hybridises with the 3' end of the denatured substrate (Figure 1C). Addition of DNA polymerase results in second-strand synthesis to produce a fully double-stranded DNA population (Figure 1D) that can then be digested with the restriction enzyme specific for the included restriction enzyme site. After digestion, this population of DNA inserts is cloned under control of the mammalian U6 snRNA promoter to produce a genome-wide shRNA expression library (Figure 1E). Upon delivery and transcription of these vectors in mammalian cells, a shRNA is produced specific for a target mRNA having complementarity with the double-stranded RNA region (Figure 1F). These shRNAs have been shown to be effective inhibitors of gene expression in mammalian cells (Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002) *Genes & Dev* 16, 948-958; Brummelkamp, T.R., Bernards, R., and Agami, R. (2002) *Science* 296, 550-553; Sui, G., Soohoo, C., Affar, E., Gay, F., Shi, Y., Forrester, W. C., and Shi, Y. (2002) *Proc Natl Acad Sci* 99, 5515-20; Yu, J., DeRuiter, S.L., and Turner, D.L. (2002) *Proc Natl Acad Sci* 99, 6047-52).

An alternative strategy to construct the DNA insert encoding the shRNA involves inclusion of a single deoxyuracil nucleotide between the primer binding site and the 19 base random sequence in the pool of single-stranded DNA initially synthesised. Prior to second strand synthesis this site is depurinated using uracil nucleotide glycosylase (UNG) and subsequently β-eliminated by boiling in piperidine. Through this process any competing component of the hairpin is removed enabling the primer to have free access to its primer binding site during second strand synthesis. Primer extension results in the production of a double stranded cassette that contains a blunt end at the depurination site and a SalI recognition sequence at the other end. As this arrangement results in the bottom strand becoming the coding strand the hairpin elbow sequence is inverted with respect to the scheme described above. This second strategy is shown Figure 5.

Example 2 p53-specific shRNA regulates p53 expression and modifies a cellular phenotype.

To demonstrate that the method described in Example 1 could be used to generate a shRNA insert encoding a shRNA capable of suppressing a specific target gene in mammalian cells, we constructed p53-specific and EGFP-specific shRNA expression plasmids. To produce the p53 shRNA insert, the following hairpin oligonucleotide is synthesised (Sigma Genosys):
5'-*GTGATTCCGTCGACC*GACTCCAGTGGTTAATCTACGTCGAGTTCAAGAGACT CGAC-3'. The sequences in italics represent the primer-binding site with the SalI site indicated by underlining. The first C following the SalI site represents the first transcribed base while the bold text indicates the sequence of the 19 nucleotides specific for the target site in p53 (Brummelkamp, T.R., Bernards, R., and Agami, R. (2002) *Science* 296, 550-553.). The stem loop structure of the hairpin oligonucleotide is composed of six complementary bases flanking the underlined loop sequence. The hairpin oligonucleotide synthesised for the EGFP target gene is
5'-*GTGATTCCGTCGACC*CGGCAAGCTGACCCTGAAGGTCGAGTTCAAGAGACT CGAC-3'. The small hairpin oligonucleotide is resuspended in 10 mM Tris pH8.8 and the concentration adjusted to 100 mM. This solution is then heated briefly (3min) to 75°C to denature any intermolecular hybrids and cooled to 37°C for 30 min. At this temperature the partial hairpin structure of the oligonucleotide is permitted to form through intramolecular hybridisation of its complementary arms. The annealed hairpin is then stored at 4°C for at least 16 hours. Then 500 pmol of the partial hairpin is transferred to a solution of 67 mM Tris-HCl (pH 8.8),16.6 mM $(NH_4)_2SO_4$, 6.7 mM $MgCl_2$, 0.8 units of Ampli*Taq* DNA polymerase (Perkin Elmer) and 300 μM each of dGTP, dATP, dTTP, dCTP and incubated a 37°C for 30 min. In some reactions a trace of [α-^{32}P]dCTP is included to enable monitoring of the reaction by subsequent electrophoresis on a 16% denaturing polyacrylamide gel. After this interval, 5 nmol of the hairpin specific primer (5'-GTGATTCCGTCGACC-3') is added (in 1Xreaction buffer) and the solution incubated at 95°C for 30 s, 50°C for 10 min and 72°C for another 10 min. The sample is then phenol extracted and ethanol precipitated in 0.3 M NaOAc and 2 volumes of ethanol before being redissolved and further purified by electrophoresed on a 2% agarose gel, band excision and recovery using the Gene Clean kit (Bio101). The purified extension product is redissolved in 10 mM Tris (pH 7.5) and digested with 10U of SalI restriction endonuclease at 37°C for 90 min then dephosphorylated by the addition of 10U of calf intestinal phosphatase. Following

second-strand synthesis, the double-stranded shRNA inserts are digested with SalI and ligated to pGUTLESS(U6+1) and pGUTLESS(U6+27), precut with SalI and XbaI and dephosphorylated, under control of the human U6 promoter. The expression vectors pGUTLESS(U6+1) and pGUTLESS(U6+27) are constructed by subcloning the U6+1 and U6+27 expression cassettes from pTZU6+1 and pAVU6+27 (Paul, C.P., Good, P.D., Winer, I, and Engelke, D.R. (2002) *Nature Biotech* 20, 505-508), respectively, as BamHI-HindIII fragments into the pd4EGFP-N1 (Clontech) previously deleted for the CMV immediate early promoter, dEGFP ORF and the SV40 polyA signal.

Each of the pGUTLESS(U6+1) and pGUTLESS(U6+27) derivative plasmids encoding the EGFP or p53-specific shRNAs can be tested in mammalian cells. For EGFP, the U6+1 and U6+27 expression vectors are co-transfected into HEK 293 cells with EGFP-N1 (target plasmid) and pSVβ (transfection control). At 48 hours post-transfection, cells are assayed for level of EGFP expression using FACs analysis. Alternatively, the EGFP shRNA-encoding vectors can be delivered to a HEK 293 cell line stably expressing the dEGFP target gene. Following selection in G418, the pooled population or selected clones can be analysed for EGFP expression as above. To test the p53-specific shRNA expression plasmids, each is linearised and transfected into HCT116 colon cancer cells. Following selection in G418 for 14 days, pooled populations are exposed to 0 μM and 200 μM 5-fluorouracil (5-FU) for 48 hours at which point p53 protein levels are examined using Western analysis.

Example 3 Screening random shRNA expression libraries.

One advantage of screening a genome-wide random RNAi expression library is the ability to identify effective targets that are linked to the change in cellular phenotype using a genetic screen. To examine the selectivity and sensitivity of such a strategy, the p53-specific shRNA expression plasmid is diluted $1:10^2$, $1:10^3$, $1:10^4$, $1:10^5$ and $1:10^6$ with vector alone. Each of these plasmid mixes is delivered to HCT116 cells as above and Neo^R cells selected in the presence of G418. Following selection, the population is then exposed to 200 μM 5-FU for 48 hrs, and clones recovered in standard growth medium. In this genetic selection, p53 wild type cells are sensitive to the chemotherapeutic drug 5-FU and are unable to undergo clonal growth after exposure. In contrast, cells showing reduced p53 are resistant and retain the ability to form clones following selection. The 5-FU resistant clones from each plasmid mix are recovered and screened for the p53-specific shRNA expression plasmid using a rapid

PCR screen. Alternatively, the p53-specific shRNA expression plasmids can be linearised and transfected into HT29 colon cancer cells stably expressing a temperature-sensitive p53 gene (McKay, B.C., Chen, F., Perumalswami, C. R., Zhang, F., and Ljungman, M. (2000) *Mol Biol Cell* 11, 2543-51). At 38 °C, these cells do not express active p53 and thus grow normally. However, when shifted to 32 °C, these cells over-express functional p53 which causes growth arrest. Thus, cells expressing an active p53-specific shRNA at 32 °C will overcome p53-induced growth arrest and form colonies at this temperature. These clones can be recovered and analysed using PCR as above.

Example 4 Generation of random siRNA expression libraries.

A second method of generating random RNAi expression libraries for genetic screening in mammalian cells is summarised in Figure 2. In this approach, a pool of single-stranded DNA is synthesised containing four adenosines, followed by a 19 base random sequence (N19), four thymidines, a restriction enzyme site (RE1), and a primer-binding site (Figure 2A). A primer is added to the pool of oligonucleotides and second-strand synthesis is directed using DNA polymerase to produce a pool of double-stranded DNAs (Figure 2B, C). This population is digested with the appropriate restriction enzyme and ligated between two convergent U6 promoters (Figure 2D), such that upon transcription of the two strands of the resident inserts, two small complementary RNAs are produced that are capable of hybridising to form a classical siRNA with two to four base overhangs at their 3′ ends (Figure 2E). These forms of siRNAs have been shown to mediate specific gene silencing in mammalian cells (Miyagishi, M. and Taira, K. (2002) *Nature Biotech* 19, 497-500; Lee, N., Dohjima, T., Bauer, G., Li, H., Li, M., Ehsani, A., Salvaterra, P., and Rossi, J. (2002) *Nature Biotech* 19, 500-505).

Example 5 p53-specific siRNA regulates p53 expression and modifies a cellular phenotype

One of the key features of the method described for constructing a random siRNA expression library is the requirement for an expression vector containing convergent U6 promoters (or alternatively, H1 or T7 promoters) controlling expression of the sense and antisense RNAs encoded by the resident inserts. Toward this end, the U6 promoter can be PCR-amplified from pGUTLESS(U6+1) using the following PCR primers: 5′-GCGCAAGCTTATAGGGAATTCGAGCTCGGTA-3′ and

5'-GCGCTCTAGAGGTGTTTCGTCCTTTCCACAA-3'. The PCR products are digested with HindIII (near the 5' end) and XbaI (near the 3' end) and cloned into pGUTLESS(U6+1), precut with XbaI and HindIII, to replace the Pol III stem terminator region (Paul, C.P., Good, P.D., Winer, I, and Engelke, D.R. (2002) *Nature Biotech* 20, 505-508). The end result is the plasmid pGUTLESS(2xU6) containing the convergent U6 promoter expression cassette.

To demonstrate the capacity of the convergent U6 promoter expression cassette to produce both sense and antisense transcripts, and mediate gene silencing in mammalian cells, EGFP- and p53-specific siRNA expression vectors are constructed. In the case of the EGFP target, the following oligonucleotides are synthesised and annealed: 5'- TCGACAAAAACGGCAAGCTGACCCTGAAGTTTTT-3' and 5'-CTAGAAAAACTTCAGGGTCAGCTTGCCGTTTTTG-3'. For the p53-specific insert, the following oligonucleotides are synthesised and annealed: 5'-TCGACAAAAAGACTCCAGTGGTTAATCTACTTTTT-3' and 5'-CTAGAAAAATAGATTAACCACTGGAGTCTTTTTG-3'. Each of these inserts is cloned into pGUTLESS(2xU6), precut with SalI and XbaI, between the convergent U6 promoters to produce the siRNA expression vectors. Each of these expression plasmids is tested as described above for the EGFP- and p53-specific shRNA plasmids.

Example 6 Construction of target gene and genome (viral, pathogen)-specific shRNA and siRNA libraries

The strategies described above allow the production of RNAi expression libraries that contain dsRNA genetic inhibitors for each of the expressed genes of any genome including mammalian cells. These same libraries also have utility for identifying both host genes and viral or pathogen-derived genes that play a major role in the susceptibility of cells to infection by viruses and pathogens. The described methods can be modified to construct RNAi expression libraries restricted to a specific viral or pathogen genome or to a limited number of targets genes. The latter application is particularly relevant for probing gene function of up- or down-regulated genes identified in large-scale microarray or subtractive hybridisation experiments where only a subset of genes is implicated in the phenotype under investigation. Figure 3 summarises the strategy for constructing target gene(s) and genome-specific shRNA and siRNA expression libraries. In the initial step, the target gene(s) or viral or pathogenic genome is treated with DNAseI to fragment the starting DNA into 19-29 bp fragments (Figure 3A). To construct a shRNA expression library, the pool of DNA

fragments is ligated to a universal hairpin sequence and all DNA fragments containing a single hairpin linker are isolated (Figure 3B). A dsDNA adaptor (containing a primer-binding site) is then ligated to the end of these DNAs (that does not contain the hairpin linker) and all fragments having a single hairpin linker and dsDNA adaptor are isolated (Figure 3C). This pool of DNA is then denatured, annealed to the universal primer, subjected to second-strand synthesis and then digested and ligated under control of the U6 promoter in a mammalian expression plasmid (Figure 3D-F). To construct a siRNA expression library, the randomly fragmented 19-29 bp DNAs are ligated to a dsDNA adaptor (as used in Figure 3C) and all DNAs containing a single set of adaptors are isolated (Figure 3G). These DNAs can either be PCR-amplified using a primer specific for the ligated adaptors (Figure 3H) or digested directly and ligated between convergent U6 promoters (Figure 3I).

Example 7 Identification of HIV therapeutics using HIV-derived shRNA libraries.

Genetic selection assays can be used to screen a HIV-specific RNAi expression library for effective RNAi construct that confer resistance to HIV infection or that interfere with the productive or latent phases of the viral life cycle. Such genetic selection assays using genetic suppressor element libraries have been described (Dunn, S.J., Park, S.W., Sharma, V., Raghu, G., Simone, J.M., Tavassoli, R., Young, L.M., Ortega, M.A., Pan, C-H., Alegre, G.J., Roninson, I.B., Lipkina, G., Dayn, A., and Holzmayer, T.A. (1999) Gene Therapy 6, 130-137) and are outlined in Figure 4. In one assay, chronically infected promyelocyctic HL60 cells, which are 99% CD4 positive until induction of latent HIV, can be induced to lose CD4 upon the addition of TNFα (type 4). Expression of an effective HIV-specific shRNA will be expected to interfere with this induction and result in the retention of CD4 on the cell surface. Cells containing effective shRNA constructs can then be separated from the CD4-negative population using FACs sorting. These constructs should be effective at inhibiting HIV induction in latently infected cells. In a second assay, CEM T4 cells infected with replicating HIV display an accumulation of p24 and a reduction of CD4 (Figure 4B). Thus, expression of an effective shRNA construct that interferes with productive infection can be identified by enriching for cells exhibiting the CD4-positive and p24-negative phenotype using FACs. Both of these genetic selection systems can identify novel HIV-specific shRNA expressing vectors that could be used as gene therapy against multiple stages of the HIV life cycle.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Dated this 4th day of September 2002

Johnson & Johnson Research Pty Ltd
Patent Attorneys for the Applicant:

BLAKE DAWSON WALDRON
PATENT SERVICES

REFERENCES

Asubel.F M et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York N.Y.

Baker, M. W., and Macagno, E. R. (2000) *Curr Biol* 10(17), 1071-4.

Baulcombe, D. C. (1996) *Plant Mol Biol* 32(1-2), 79-88.

Beaucage and Caruthers (1981) Tetrahedron Letts. 22(20):1859-1862

Billy, E., Brondani, V., Zhang, H., Muller, U. and Filipowicz, W. (2001) *Proc. Natl Acad Sci* 98, 14428-14483.

Brummelkamp, T.R., Bernards, R., and Agami, R. (2002) *Science* 296, 550-553.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001) *Proc. Natl. Acad. Sci.* 98, 9742-9747.

Cogoni, C., and Macino, G. (1999) *Nature* 399 (6732), 166-9.

Dunn, S.J., Park, S.W., Sharma, V., Raghu, G., Simone, J.M., Tavassoli, R., Young, L.M., Ortega, M.A., Pan, C-H., Alegre, G.J., Roninson, I.B., Lipkina, G., Dayn, A., and Holzmayer, T.A. (1999) *Gene Therapy* 6, 130-137.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001) *Genes & Dev* 15:188-200.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001) *Nature* 411(6836), 494-8.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) *Nature* 391(6669), 806-11.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000) *Nature* 404: 293-96

Hammond, S.M., Caudy, A.A., and Hannon, G.J. (2001) *Nat. Rev. Genet.* 2, 110-19.

Hofacker et al. (1994) Monatshefte F. Chemie 125:167-188

Hsu, J.Y., Reimann, J. D. R., Sorensen, C.S., Lucas, J., and Jackson, P. K. (2002) *Nature Cell Biol.* 4, 358-366.

Kennerdell, J. R., and Carthew, R. W. (1998) *Cell* 95(7), 1017-26.

Kumar, M., and Carmichael, G. G. (1998) *Microbiol Mol Biol Rev* 62(4), 1415-34.

Lee, N., Dohjima, T., Bauer, G., Li, H., Li, M., Ehsani, A., Salvaterra, P., and Rossi, J. (2002) *Nature Biotech* 19, 500-505.

Lohmann, J. U., Endl, I., and Bosch, T. C. (1999) *Dev Biol* 214(1), 211-4.

Maxam and Gilbert (1980) in Grossman and Moldave (eds.) Academic Press, New York, Methods in Enzymology 65:499-560.

McCaskill (1990) Biopolymers 29:1105-1119.

McKay, B.C., Chen, F., Perumalswami, C. R., Zhang, F., and Ljungman, M. (2000) *Mol Biol Cell* 11, 2543-51.

Miyagishi, M. and Taira, K. (2002) *Nature Biotech* 19, 497-500.

Needham-VanDevanter et al (1984) Nucleic Acids Res., 12:6159-6168.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998) *Proc Natl Acad Sci USA* 95(25), 14687-92.

Paddison, P., Caudy, A. A., and Hannon, G.J. (2002) *Proc. Natl Acad. Sci.* 99, 1443-1448.

Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002) *Genes & Dev* 16, 948-958.

Paul, C.P., Good, P.D., Winer, I, and Engelke, D.R. (2002) *Nature Biotech* 20, 505-508.

Pearson and Regnier (1983) J. Chrom. 255:137-149.

Sambrook et al (1989) Molecular Cloning, A laboratory Manual, Cold Spring Harbor Press, Plainview N.Y.

Schoppmeier, M., and Damen, W. G. (2001) *Dev Genes Evol* 211(2), 76-82.

Sui, G., Soohoo, C., Affar, E., Gay, F., Shi, Y., Forrester, W. C., and Shi, Y. (2002) *Proc Natl Acad Sci* 99, 5515-20.

Svoboda, P., Stein, P., Hayashi, H., and Schultz, R. M. (2000) *Development* 127(19), 4147-4156.

Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A., and Driscoll, M. (2000) *Nat Genet* 24(2), 180-3.

Thompson, B., Tonwsley, F., Rosin-Arbesfeld, R., Muisi, H., and Bienz, M. (2002) *Nature Cell Biol.* 4, 367-373.

Wargelius, A., Ellingsen, S., and Fjose, A. (1999) *Biochem Biophys Res Commun* 263(1), 156-61.

Wianny, F., and Zernicka-Goetz, M. (2000) *Nat Cell Biol* 2(2), 70-5.

Yang, S., Tutton, S., Pierce, E., and Yoon, K. (2001) *Mol Cell Biol* 21(22), 7807-16.

Yu, J., DeRuiter, S.L., and Turner, D.L. (2002) *Proc Natl Acad Sci* 99, 6047-52.

Figure 1. Generation of a random shRNA library

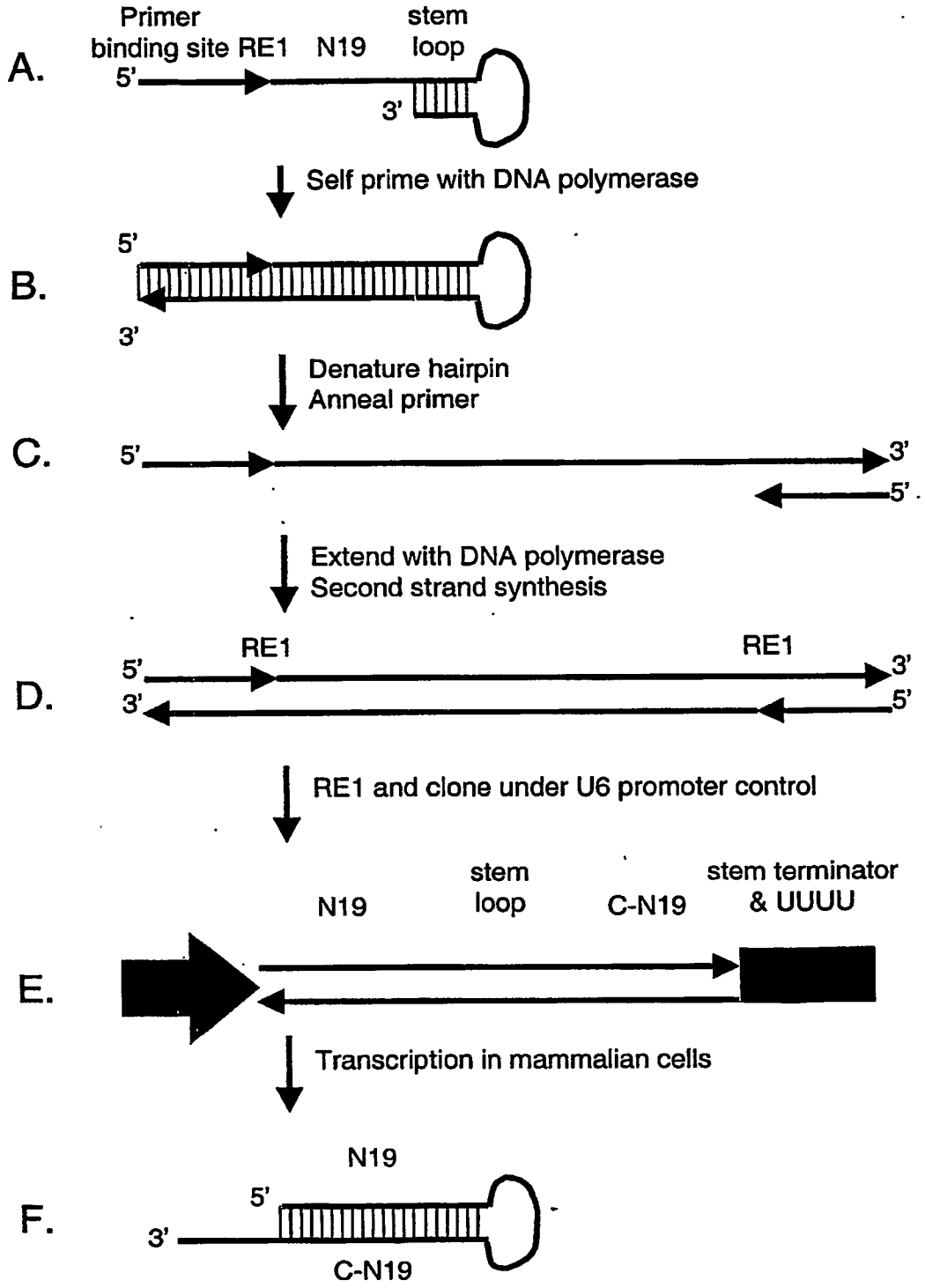

Figure 2. Generation of a random siRNA library

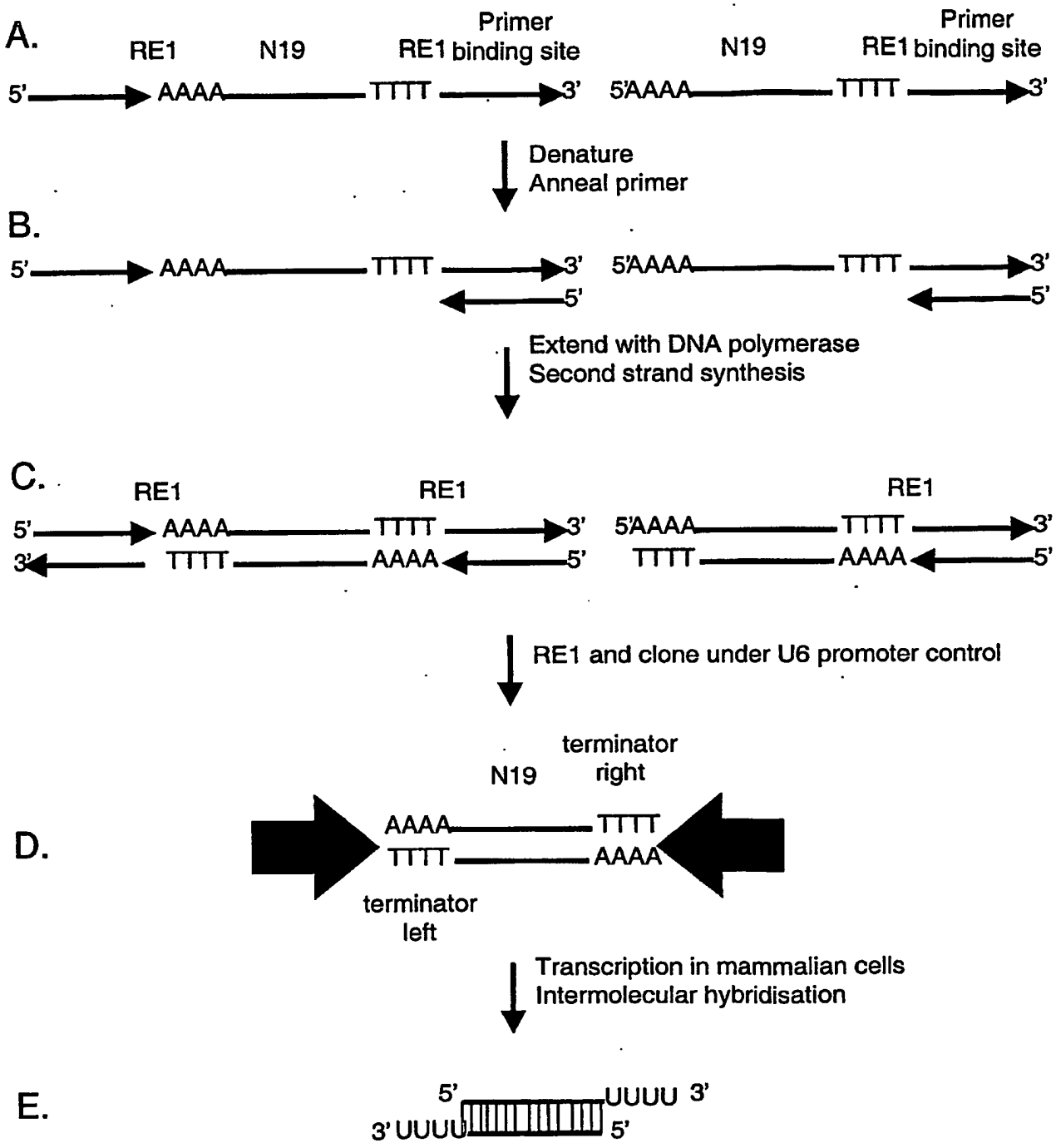

Figure 3. Method for constructing genome-specific shRNA and siRNA libraries

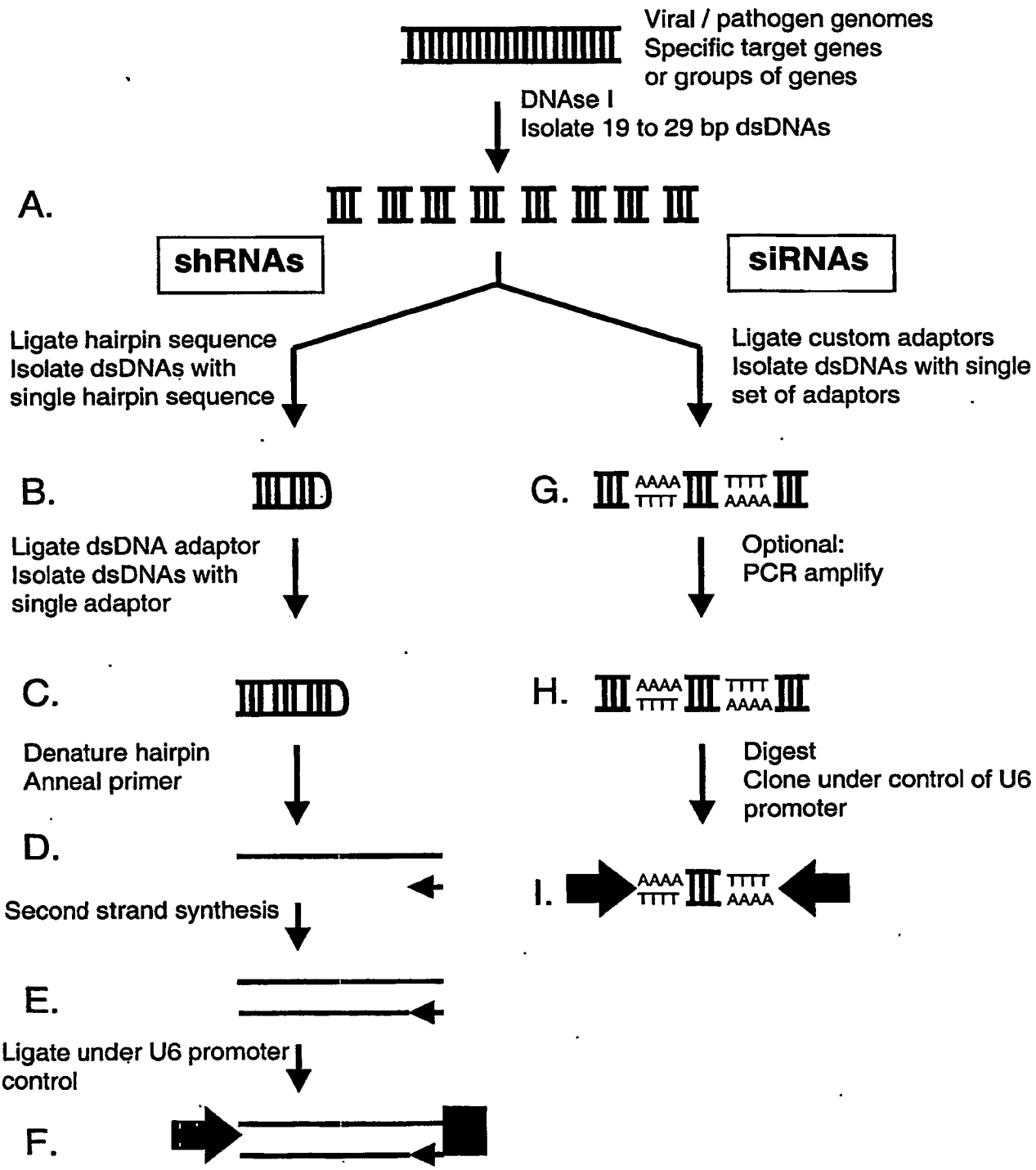

Figure 4. Identification of HIV-specific shRNA or siRNAs using genetic selections

A. Genetic selection for HIV-specific shRNAs blocking induction of HIV

Transfect universal RNAi retroviral expression library into producer cells

↓

Transduce chronically infected promyelocytic clone of HL60

↓

Induce with TNF-α and sort for CD4 positive cells

↓

Purify genomic DNA, PCR amplify shRNA inserts and sequence

↓

Re-test independent shRNA constructs

B. Genetic selection for HIV-specific shRNAs blocking productive infection by HIV

Transfect universal RNAi retroviral expression library into producer cells

↓

Transfect CEM T4 cells

↓

Infect with HIV-1IIIB and sort for CD4 positive and p24 negative cells

↓

Purify genomic DNA, PCR amplify shRNA inserts and sequence

↓

Re-test independent shRNA constructs

Figure 5

1) p53 ShRNA expression cassette with 6base stem and a 9base loop (after Brummelkamp et al., 2002).

GACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGAC

2) Add primer binding site

TGTGGTGATTCCGTCGACUGACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGAC

3) Synthesize hairpin oligonucleotide and primer

TGTGGTGATTCCGTCGACUGACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGAC

4) Extend self primer with DNA polymerase

TGTGGTGATTCCGTCGACUGACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGACGTAGATTACCACTGGAGTCAGTCGACGGAATCACCACA

5) Treat with UNG and boil in piperidine, anneal primer and

GACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGACGTAGATTACCACTGGAGTCAGTCGACGGAATCACCACA

GCTGCCTTAGTGGTGT-5´

extend with DNA polymerase

GACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGACGTAGATTACCACTGGAGTCAGTCGACGGAATCACCACA

CTGAGGTCACCATTAGATGCAGCTCAGAGAACTTGAGCTGCATCTAATGGTGACCTCAGTCAGCTGCCTTAGTGGTGT-5´

6) Digest with SalI & Dephosphorylate

GACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGACGTAGATTACCACTGGAGTCAG

CTGAGGTCACCATTAGATGCAGCTCAGAGAACTTGAGCTGCATCTAATGGTGACCTCAGTCAGCT-5´

7) Ligate to SalI/Xba cut vector

```
..gct          GACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGACGTAGATTACCACTGGAGTCAG          tcgacggtgt
..cgagatc      CTGAGGTCACCATTAGATGCAGCTCAGAGAACTTGAGCTGCATCTAATGGTGACCTCAGTCAGCT-5´      gccaca

..gct        GACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGACGTAGATTACCACTGGAGTCAGtcgacggtgt
..cgagatc    CTGAGGTCACCATTAGATGCAGCTCAGAGAACTTGAGCTGCATCTAATGGTGACCTCAGTCAGCTgccaca
```

8) End fill the protruding single stranded end of vector with DNA polymerase.

..gctctag GACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGACGTAGATTACCACTGGAGTCAGtcgacggtgt

..cgagatc CTGAGGTCACCATTAGATGCAGCTCAGAGAACTTGAGCTGCATCTAATGGTGACCTCAGTCAGCTgccaca

9) Blunt end ligate to circularise recombinant

..aaaagcggaccgaagtccgctctagGACTCCAGTGGTAATCTACGTCGAGTCTCTTGAACTCGACGTAGATTACCACTGGAGTCAGtcgacggtgt

..ttttcgcctggcttcaggcgagatcCTGAGGTCACCATTAGATGCAGCTCAGAGAACTTGAGCTGCATCTAATGGTGACCTCAGTCAGCTgccaca

10) Transcript produced in the U6+1 construct

GUCGACUGACUCCAGUGGUAAUCUACGUCGAGUUCAAGAGACUCGACGUAGAUUACCACUGGAGUCCUAGAGCGGACUUCGGUCCGCUUUU